

SHORT COMMUNICATION

Lowering sample size in comparative analyses can indicate a correlation where there is none: example from Rensch's rule in primates

P. LINDENFORS* & B. S. TULLBERG†

University of Virginia, Department of Biology, Charlottesville, VA, USA†Department of Zoology, Stockholm University, Stockholm, Sweden**Keywords:*

allometry;
 phylogenetic comparative analyses;
 sexual size dimorphism.

Abstract

The fact that characters may co-vary in organism groups because of shared ancestry and not always because of functional correlations was the initial rationale for developing phylogenetic comparative methods. Here we point out a case where similarity due to shared ancestry can produce an undesired effect when conducting an independent contrasts analysis. Under special circumstances, using a low sample size will produce results indicating an evolutionary correlation between characters where an analysis of the same pattern utilizing a larger sample size will show that this correlation does not exist. This is the opposite effect of increased sample size to that expected; normally an increased sample size increases the chance of finding a correlation. The situation where the problem occurs is when co-variation between the two continuous characters analysed is clumped in clades; e.g. when some phylogenetically conservative factors affect both characters simultaneously. In such a case, the correlation between the two characters becomes contingent on the number of clades sharing this conservative factor that are included in the analysis, in relation to the number of species contained within these clades. Removing species scattered evenly over the phylogeny will in this case remove the exact variation that diffuses the evolutionary correlation between the two characters – the variation contained within the clades sharing the conservative factor. We exemplify this problem by discussing a parallel in nature where the described problem may be of importance. This concerns the question of the presence or absence of Rensch's rule in primates.

Introduction

The *raison d'être* for phylogenetic comparative methods is that characters sometimes co-occur in different species because of common descent and not always because of a functional relationship between the characters in question. A familiar example of this is the co-occurrence of mammary glands and fur. These are shared by all mammals, but the functional relationship is – as far as we know – nonexistent; we could as well have been feathered.

A simple example of the problems such issues can cause was given by Felsenstein (1985) in the context of

the initial description of the independent contrasts method. This hypothetical example consisted of a correlation found between two continuous characters, where it turned out that the data was clumped due to common descent, so that one group of related species had high values in both the *x*- and *y*-variables while another group had low values in both variables. The correlation *within* these two groups of related species was nonexistent (Fig. 1a). The correlation found between the variables was therefore due to a single event where the ancestor of one or the other group had undergone a unique change in both variables – a situation that independent contrasts can handle.

There exist more fine-grained degrees of this problem, however, where it is not always even obvious that there actually exists a problem. This concerns similar situations

Correspondence: P. Lindenfors, University of Virginia, Department of Biology, Gilmer Hall, Charlottesville, VA 22903, USA.
 Tel.: 254-735-502 811; fax: +46 816 7715;
 e-mail: Patrik.Lindenfors@virginia.edu

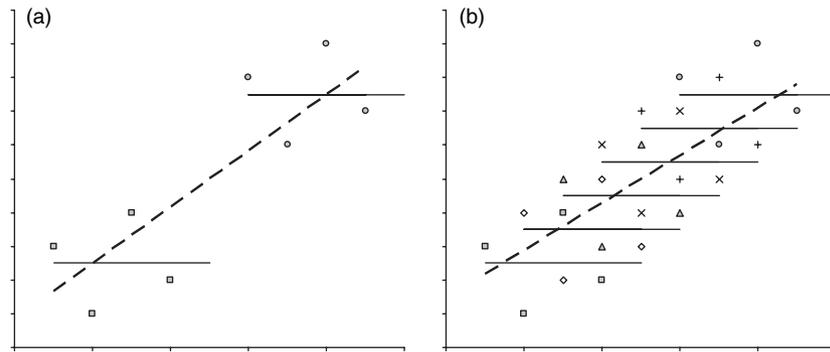


Fig. 1 A generalization of Felsenstein (1985) example, where a correlation is found because of correlations existing *between* monophyletic groups (dashed lines), but not *within* them (solid lines). In case (a) two monophyletic groups happen to differ from each other in both the x - and y -variable, so that when the groups are analysed disregarding phylogeny, a correlation is found. In case (b) the same two groups are now accompanied by four more showing the exact same pattern; the correlation between the x - and y -variable only exists between monophyletic groups, while the correlation is non-existent within groups. Disregarding – or not knowing – phylogenetic relationships or the biology of the organisms investigated thus means that the ‘true’ relationship remains hidden. As an extension, it is interesting to note that the rate of extinction and speciation under such circumstances can influence the results of comparative analyses.

to that exemplified by Felsenstein (1985); a potential correlation between two continuous characters where the distributions of the x - and y -variables are clumped due to common descent, but this time in several clades instead of just two (Fig. 1b). If these clades are numerous enough in relation to the species contained within them, a correlation may be found without the clumping being obvious enough to be detected. In such a scenario, the number of data points becomes crucial – but in the opposite manner to that normally expected.

Any basic statistics text will inform you that increasing sample size will increase the probability of detecting existing significant trends in the data (e.g. Sokal & Rohlf, 1995). Therefore, the notion that there exist cases where the opposite is true may be counterintuitive – i.e. cases where a decreased sample size increases the probability of detecting trends in the data (i.e. committing a Type-I error). This note aims at pointing out conditions when this will occur. Basically, the cause is the same as that Felsenstein (1985) highlighted – similarity due to shared descent rather than a functional correlation – but replicated many times in the phylogeny (Fig. 1). We also want to highlight the importance of considering evolutionary causes when analyzing evolutionary patterns. For the purpose of clarity we follow Felsenstein's (1985) method of illustration by employing an idealized example. We also provide a corresponding parallel from nature where the problem may appear – this concerns the existence of Rensch's rule in primates.

Rensch's rule describes the pattern that sexual size dimorphism tends to scale with body size in related species (Abouheif & Fairbairn, 1997; Fairbairn, 1997) – a rule that takes its name from the researcher who first described it (Rensch, 1950, 1959). A number of analyses of the presence/absence of this pattern have been carried

out in primates, most finding that primates clearly exhibit Rensch's rule (Ralls, 1977; Clutton-Brock *et al.*, 1977; Leutenegger, 1978; Leutenegger & Cheverud, 1982; Gaulin & Sailer, 1984; Reiss, 1986; Abouheif & Fairbairn, 1997; Smith & Cheverud, 2002), with the exception of strepsirrhines (Kappeler, 1990), and with a weaker trend in platyrrhines (Ford, 1993). We have, however, reported earlier that support for this pattern disappears if enough data are included and phylogeny is controlled for (Lindenfors & Tullberg, 1998).

All the above-presented studies of primate dimorphism used different phylogenies when checking for the presence of Rensch's rule, or – indeed – did not correct for phylogeny at all. Although this is highly likely to have influenced the results, it is not meaningful to argue for one phylogeny over the other here. We instead want to emphasize another difference between these studies: that we (Lindenfors & Tullberg, 1998) utilized a larger sample size than did any of the other studies. By conventional logic, however, this would mean that we would be more likely to detect the presence of Rensch's rule, not less likely (Sokal & Rohlf, 1995). Note that there may be important reasons for opting for a low sample size. Smith & Cheverud (2002); $n = 105$, for example, excluded a number of species present in Lindenfors & Tullberg (1998); $n = 147$ and in Smith & Jungers (1997); $n = 230$; they excluded species where weight data were gathered from less than five individuals per species and sex – a quick and simple quality check on the data. In this specific case, however, this quality check introduces another error source; one related to the probable causes behind Rensch's rule in primates.

Empirical studies indicate that sexual size dimorphism in primates is primarily caused by intrasexual selection on male size (e.g. Alexander *et al.*, 1979; Clutton-Brock &

Harvey, 1977; Gaulin & Sailer, 1984; Harvey & Harcourt, 1984; Mitani *et al.*, 1996; Lindenfors & Tullberg, 1998). Theoretical work suggests that female size also should increase even if the selection on body size is male-specific due to a genetic correlation between the sexes (Maynard Smith, 1978; Lande, 1985, 1987; Lande & Arnold, 1983), but this is expected to be only a temporary phenomenon; female size at equilibrium should be independent of male size (Lande, 1985; Fairbairn, 1997; Reeve and Fairbairn, 2001). There are, however, reasons to expect selection on female size that is correlated with degree of sexual size dimorphism. Female primates may simply need to be larger in species where males are larger because the females are the ones raising these larger males (Fairbairn, 1997; Lindenfors, 2002). There are, for example, results showing that males in dimorphic primate species are larger already at birth (Smith & Leigh, 1998; but see Lindenfors, 2002) and that age at weaning is significantly higher in sexually selected species, even after correcting for body mass (Lindenfors, 2002). Especially age at weaning is important, because resources transferred during lactation are typically more energetically costly than the prenatal costs of gestation (Cameron, 1998) and it is also a general pattern in mammals that female body mass is correlated to milk yield (Oftedal, 1984). Therefore, a longer suckling period constitutes a significant selection pressure on female size. Sexual selection on males and correlated selection on females are thus the causes indicated by empirical studies to lie behind the correlation between body size and body size dimorphism – significant or not – in primates (Lindenfors, 2002).

If sexual selection on males with a correlated response in females is – as the evidence suggests – the process behind Rensch's rule, however, then a low number of data points – with omitted species being evenly scattered over the phylogeny – increases the probability of making a Type-I error. To show why this is so and to illustrate the problem clearly, a highly idealized example is provided (Fig. 2). To highlight the parallel with Rensch's rule in primates, the variables are named accordingly, but the example concerns any situation where one variable – known or unknown – is correlated with the two variables being analysed.

A measurement of sexual selection that is often used in studies on primates is mating system. Note that this measure varies between on (polygyny) and off (monogamy) in this example. The degree of sexual selection in primates of course varies on a finer level than this (e.g. Mitani *et al.*, 1996), but the general reasoning is the same for a continuous measurement as for a categorical variable. The only prerequisite for the error we want to highlight is that the degree of sexual selection is clumped in the phylogeny. In primates there are plenty of examples of this; for example, all macaques are reported as being multi-male multifemale and all gibbons as being monogamous (Smuts *et al.*, 1987). It follows that if sexual selection is the process behind Rensch's rule, then the

body size of females and body size dimorphism should co-vary only in comparisons *between* clades differing in mating system, while comparisons *within* a clade with a common mating system would show no or little such co-variation.

In the example given here, body size – and to some degree also body size dimorphism – has varied randomly during evolution. In addition, however, sexual selection has acted on male body size so that more polygynous species have larger males and larger dimorphism than less polygynous species. Furthermore, more polygynous species also have larger females than less polygynous species. This pattern matches that empirically found in primates (Abouheif & Fairbairn, 1997; Lindenfors & Tullberg, 1998; Lindenfors, 2002; Thorén *et al.* in press).

We here adhere to the method choice of Abouheif & Fairbairn (1997), carrying out an independent contrasts analysis, regressing male body size contrasts on to female body size contrasts and then testing for a difference with a slope of 1.0. The alternative method of regressing size dimorphism on female size is to be avoided for statistical reasons (e.g. Atchley *et al.*, 1976; Ranta *et al.*, 1994; Sokal & Rohlf, 1995; but see Smith, 1999) and because male and female body sizes are what selection acts upon – not dimorphism *per se*. Furthermore, we use major axis regressions through the origin as there is no a priori reason to put males or females on the *x*- or *y*-axis. Body mass was \log_{10} -transformed prior to analysis. The analyses were carried out using independent contrasts as implemented in the computer program PDAP (Garland *et al.*, 1993) and unit branch lengths.

Our analysis of this example reveals that Rensch's rule is unsupported by the data (regression through the origin $b = 1.092$, $R^2 = 0.892$, $n = 53$, $P = 0.152$; Fig. 3a). On the other hand, if one deletes a subset of the data evenly distributed over the phylogeny (~33%, Fig. 2) and repeats the analysis, an independent contrasts analysis now provides significant support for Rensch's rule (regression through the origin $b = 1.140$, $R^2 = 0.926$, $n = 35$, $P = 0.023$; Fig. 3b).

As the example illustrates, by removing species evenly scattered over the phylogeny one removes exactly the variation that would diffuse the relationship between female body size and body size dimorphism – the variation within clades sharing the same mating system. This type of problem is the initial rationale behind phylogenetic comparative methods (Felsenstein, 1985): closely related species may share characteristics due to descent. Thus, by removing data-points that are insecure, another source of error is introduced.

As pointed out earlier, this problem will appear in any situation where one variable is correlated with the two variables being analysed. There is, of course, a scale of this problem from the situation that Felsenstein (1985) described where the species were clumped in two distinct clades and the correlation between the two variables was

Species	Female body size	Male body size	Ratio male/female	Mating system
Species 01	11.00	12.10	1.10	PG
Species 02	10.00	11.50	1.15	PG
Species 03	9.00	9.45	1.05	PG
Species 04	9.00	8.55	0.95	Mon
Species 05	8.00	8.40	1.05	Mon
Species 06	7.00	7.00	1.00	Mon
Species 07	10.00	11.00	1.10	PG
Species 08	9.00	10.35	1.15	PG
Species 09	8.00	8.40	1.05	PG
Species 10	8.00	7.60	0.95	Mon
Species 11	7.00	7.35	1.05	Mon
Species 12	6.00	6.00	1.00	Mon
Species 13	9.00	9.90	1.10	PG
Species 14	8.00	9.20	1.15	PG
Species 15	7.00	7.35	1.05	PG
Species 16	7.00	6.65	0.95	Mon
Species 17	6.00	6.30	1.05	Mon
Species 18	5.00	5.00	1.00	Mon
Species 19	11.00	12.10	1.10	PG
Species 20	10.00	11.50	1.15	PG
Species 21	9.00	9.45	1.05	PG
Species 22	9.00	8.55	0.95	Mon
Species 23	8.00	8.40	1.05	Mon
Species 24	7.00	7.00	1.00	Mon
Species 25	10.00	11.00	1.10	PG
Species 26	9.00	10.35	1.15	PG
Species 27	8.00	8.40	1.05	PG
Species 28	8.00	7.60	0.95	Mon
Species 29	7.00	7.35	1.05	Mon
Species 30	6.00	6.00	1.00	Mon
Species 31	9.00	9.90	1.10	PG
Species 32	8.00	9.20	1.15	PG
Species 33	7.00	7.35	1.05	PG
Species 34	7.00	6.65	0.95	Mon
Species 35	6.00	6.30	1.05	Mon
Species 36	5.00	5.00	1.00	Mon
Species 37	11.00	12.10	1.10	PG
Species 38	10.00	11.50	1.15	PG
Species 39	9.00	9.45	1.05	PG
Species 40	9.00	8.55	0.95	Mon
Species 41	8.00	8.40	1.05	Mon
Species 42	7.00	7.00	1.00	Mon
Species 43	10.00	11.00	1.20	PG
Species 44	9.00	10.35	1.30	PG
Species 45	8.00	8.40	1.10	PG
Species 46	8.00	7.60	0.95	Mon
Species 47	7.00	7.35	1.05	Mon
Species 48	6.00	6.00	1.00	Mon
Species 49	9.00	9.90	1.20	PG
Species 50	8.00	9.20	1.30	PG
Species 51	7.00	7.35	1.10	PG
Species 52	7.00	6.65	0.95	Mon
Species 53	6.00	6.30	1.05	Mon
Species 54	5.00	5.00	1.00	Mon

Fig. 2 Sample phylogeny and data. For sake of illustration the phylogeny is small and homogenous and the data chosen to match the pattern empirically found in primates. Thus, larger body size and body size dimorphism is not distributed evenly over the phylogeny, but clumped in clades and related to the presence of polygyny. Species indicated with a strikethrough are those omitted for the second analysis. These are evenly spread over the phylogeny. See text for further details.

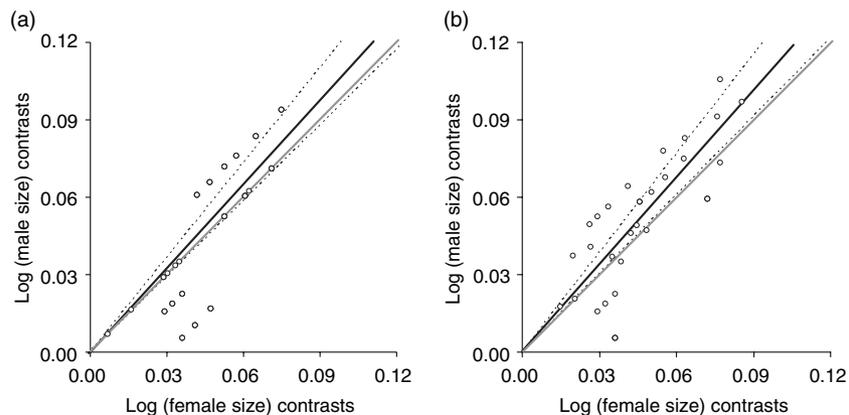


Fig. 3 The relationship between independent contrasts in male and female body size for (a) the full data set ($n = 53$) and (b) the truncated data set ($n = 35$). One data point in the graphs represents an independent contrast, the thick black lines represents the major axis regression slopes while the dashed lines show the 95% confidence intervals, which overlap with a slope of 1.0 (grey lines) in the full data set, but not in the truncated data set. Thus, in case the cause for the correlation is phylogenetically conservative (i.e. gathered in clades), then limiting the number of data points actually increases the probability of finding a correlation. (It may appear that (a) has more data points than (b) and that neither have the number of contrasts specified. This is because many contrasts are identical which in turn is due to the symmetrical construction of the example concerning both tree-shape and variable values. This symmetry is not a prerequisite for the problem outlined here, but is instead aimed at simplifying comprehension).

plainly 'false', to the opposite end of the spectrum where the 'clumps' consist of single species and a correlation thus will be plainly 'true'. However, if potential evolutionary causes or correlates are hidden, unknown or disregarded, such an underlying pattern will never be exposed. Unsurprisingly, no choice of method – not even the independent contrasts method that was made for this type of situation – can replace biological knowledge. Interestingly, since the probability of picking up the type of relationship described here depends on the number of species within groups in proportion to the number of groups, extinctions evenly spread over a phylogeny will hide relationships such as these.

We hope that we have highlighted the futility in the quest for the 'truth' about Rensch's rule in primates. Though the work to explain why body size and size dimorphism covaries in a number of animal groups has greatly furthered our understanding of evolutionary processes (e.g. Fairbairn, 1997), this understanding does not hinge on whether Rensch's rule is significantly present or not. The difference between a significant and a non-significant result may only be due to including a few more or a few less species, but the evolutionary processes remain the same regardless.

Acknowledgments

We wish to thank Daphne Fairbairn, Theodore Garland Jr., Niklas Janz and one anonymous reviewer for comments on previous drafts of this manuscript. This study was supported by The Swedish Research Council (to PL and grant no. B 621-2001-2981 to BT).

References

- Abouheif, E. & Fairbairn, D.J. 1997. A comparative analysis of allometry for sexual size dimorphism: assessing Rensch's rule. *The American Naturalist* **149**: 540–562.
- Alexander, R.D., Hoogland, J.L., Howard, R.D., Noonan, K.M. & Sherman, P.W. 1979. Sexual dimorphism and breeding systems in pinnipeds, ungulates, primates and humans. In: *Evolutionary Biology and Human Social Behaviors: an Anthropological Perspective*. (N. Chagnon, & W. Irons, eds), pp. 402–435. North Scituate Massachusetts, Duxbury.
- Atchley, W.R., Gaskins, C.T. & Andersson, D. 1976. Statistical properties of ratios. I. empirical results. *Systematic Zoology* **25**: 137–148.
- Cameron, E. Z. 1998. Is suckling behaviour a useful predictor of milk intake? A review. *Animal Behaviour* **56**: 521–532.
- Clutton-Brock, T.H. & Harvey, P.H. 1977. Primate ecology and social organisation. *Journal of Zoology* **183**: 1–39.
- Clutton-Brock, T.H., Harvey, P.H. & Rudder, B. 1977. Sexual dimorphism, socioeconomic sex ratio and body weight in primates. *Nature* **269**: 797–800.
- Fairbairn, D.J. 1997. Allometry for sexual size dimorphism: pattern and process in the coevolution of body size in males and females. *Annual Review of Ecology and Systematics* **28**: 659–687.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *The American Naturalist* **125**: 1–15.
- Ford, S.M. 1993. Evolution of sexual dimorphism in body weight in platyrrhines. *American Journal of Primatology* **34**: 221–244.
- Garland, T. Jr., Dickerman, A.W., Janis, C.M. & Jones, J.A. 1993. Phylogenetic analysis of covariance by computer simulation. *Systematic Biology* **42**: 265–292.
- Gaulin, S.J.C. & Sailer, L.D. 1984. Sexual dimorphism in weight among the primates: the relative impact of allometry and

- sexual selection. *International Journal of Primatology* **5**: 515–535.
- Harvey, P.H. & Harcourt, A.H. 1984. Sperm competition, testes size, and breeding system in primates. In: *Sperm Competition and the Evolution of Animal Mating Systems*. (R. L. Smith, ed), pp. 589–659. Academic Press, London.
- Kappeler, P.M. 1990. The evolution of sexual size dimorphism in prosimian primates. *American Journal of Primatology* **21**: 201–214.
- Lande, R. 1985. Genetic and evolutionary aspects of allometry. In: *Size and Scaling in Primate Biology*. (W. L. Jungers ed), pp. 21–32. Plenum Press, London.
- Lande, R. 1987. Genetic correlations between the sexes in the evolution of sexual dimorphism and mating preferences. In: *Sexual Selection: Testing the Alternatives*. (J. W. Bradbury, & M. B. Anderson, eds.), pp. 83–94. Chichester, Wiley.
- Lande, R. & Arnold, S.J. 1983. The measurement of selection on correlated characters. *Evolution* **37**: 1210–1226.
- Leutenegger, W. 1978. Scaling of sexual dimorphism in body size and breeding system in primates. *Nature* **272**: 610–611.
- Leutenegger, W. & Cheverud, J. 1982. Correlates of sexual dimorphism in primates: Ecological and size variables. *International Journal of Primatology* **3**: 387–402.
- Lindenfors, P. 2002. Sexually antagonistic selection on primate size. *Journal of Evolutionary Biology* **15**: 595–607.
- Lindenfors, P. & Tullberg, B.S. 1998. Phylogenetic analyses of primate size evolution: the consequences of sexual selection. *Biological Journal of the Linnean Society* **64**: 413–447.
- Maynard Smith, J. 1978. *The Evolution of Sex*. Cambridge University Press, Cambridge.
- Mitani, J.C., Gros-Louis, J. & Richards, A.F. 1996. Sexual dimorphism, the operational sex ratio, and the intensity of male competition in polygynous primates. *The American Naturalist* **147**: 966–980.
- Oftedal, O.T. 1984. Milk composition, milk yield and energy output at peak lactation: a comparative review. *Symposia of the Zoology Society of London* **51**: 33–85.
- Ralls, K. 1977. Sexual dimorphism in mammals: avian models and unanswered questions. *The American Naturalist* **111**: 917–938.
- Ranta, E., Laurila, A. & Elmer, J. 1994. Reinventing the wheel: analysis of sexual dimorphism in body size. *Oikos* **70**: 313–321.
- Reeve, J.P. & Fairbairn, D.J. 2001. Predicting the evolution of sexual size dimorphism. *Journal of Evolutionary Biology* **14**: 244–254.
- Reiss, M.J. 1986. Sexual dimorphism in body size: Are larger species more dimorphic? *Journal of Theoretical Biology* **121**: 163–172.
- Rensch, B. 1950. Die Abhängigkeit der relativen Sexualdifferenz von der Körpergröße. *Bonner Zoologische Beiträge* **1**: 58–69.
- Rensch, B. 1959. *Evolution Above the Species Level*. Methuen and Co. Ltd., London.
- Smith, R.J. 1999. Statistics of sexual size dimorphism. *Journal of Human Evolution* **36**: 423–459.
- Smith, R.J. & Cheverud, J.M. 2002. Scaling of sexual dimorphism in body mass: a phylogenetic analysis of Rensch's rule in primates. *International Journal of Primatology* **23**: 1095–1135.
- Smith, R.J. & Jungers, W.L. 1997. Body mass in comparative primatology. *Journal of Human Evolution* **32**: 523–559.
- Smith, R.J. & Leigh, S.R. 1998. Sexual dimorphism in primate neonatal body mass. *Journal of Human Evolution* **34**: 173–201.
- Smuts, B.B., Cheney, D.L., Seyfarth, R.M., Wrangham, R.W. & Struhsaker, T.T. (eds) 1987. *Primate Societies*. University of Chicago Press, Chicago.
- Sokal, R.R. & Rohlf, F.J. 1995. *Biometry*, 3rd edn. W. H. Freeman and Company, New York.
- Thorén, S., Lindenfors, P. & Kappeler, P.M. 2006. Phylogenetic analyses of dimorphism in primates: evidence for stronger selection on canine size than on body size. *The American Journal of Physical Anthropology*. In press.

Received 28 November 2005; revised 9 December 2005; accepted 12 December 2005